CONTENTS

4 Point Estimation

Definition 1. A statistic, $T = t(X_1, X_2, \ldots, X_n)$, that is used to estimate the value of $\tau(\theta)$ is called an **estimator** of $\tau(\theta)$, and an observed value of the statistic, $\mathbf{t} = t(x_1, x_2, \ldots, x_n)$, is called an **estimate** of $\tau(\theta)$.

4.1 Method of Moments

Definition 2. Consider a population pdf, $f(x; \theta_1, \theta_2, \ldots, \theta_k)$, depending on one or more parameters $\theta_1, \theta_2, \ldots, \theta_k$. The moments about the origin (raw moments)are

$$\mu'_j(\theta_1, \theta_2, \ldots, \theta_k) = E(X^j), j = 1, 2, \ldots, k$$

Definition 3. If $X_1, X_2, \ldots, X_n$ is a random sample from $f(x; \theta_1, \theta_2, \ldots, \theta_k)$, the first k sample moments are given by

$$M'_j = \frac{\sum_{i=1}^{n} X_i^j}{n}, j = 1, 2, \ldots, k$$

Definition 4. The method of moments principle is to choose as estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_k$ the values $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_k$ that render the population moments equal to the sample moments. In other words, $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_k$ are solutions of the equations

$$M'_j = \mu'_j(\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_k), j = 1, 2, \ldots, k$$

Example 1. Consider a random sample from a distribution with two unknown parameters, the mean μ and the variance σ^2. Find the MMEs of μ and σ^2.

Example 2. Consider a random sample from a two-parameter exponential distribution, $X \sim EXP(1, \eta)$.

Example 3. Consider a random sample from a gamma distribution, $X \sim GAM(\alpha, \theta)$. Find the MMEs of α and θ.

Example 4.

Consider $f(x|p) = \begin{cases} \alpha, & x = 0 \\ (1-\alpha)\binom{9}{x} p^x (1-p)^{9-x}, & x = 0, 1, 2, \ldots, 9. \\ 0, & \text{otherwise} \end{cases}$

Suppose parameters are $\alpha \in [0, 1]$ and $p \in [0.1]$. Then, for $X_1, X_2, \ldots, X_n$ iid with this distribution, find a method of moments estimator for the parameter vector (α, p) based on the first two sample moments.

4.2 Method of Maximum Likelihood

Definition 5. Likelihood Function The joint density function of n random variables $X_1, \ldots, , X_n$ evaluated at $x_1, \ldots, x_n$, say $f(x_1, \ldots, x_n; \theta)$, is referred to as a likelihood function. For fixed $x_1, \ldots, x_n$ the likelihood function is a function of θ and often is denoted by $L(\theta)$. If $X_1, \ldots, , X_n$ represents a random sample from $f(x_1, \ldots, x_n; \theta)$, then

$$L(\theta) = f(x_1; \theta) \cdots f(x_n; \theta)$$

Definition 6. Maximum Likelihood Estimator Let $L(\theta) = f(x_1, \ldots, x_n; \theta)$, be the joint pdf of $X_1, \ldots, , X_n$. For a given set of observations, $(x_1, \ldots, x_n)$, a value $\hat{\theta}$ in Ω at which $L(\theta)$ is a maximum is called a **maximum likehhood estimate** (MLE) of θ. That is $\hat{\theta}$ is a value of θ that satisfies

$$f(x_1, \ldots, x_n, ; \hat{\theta}) = \max_{\theta \in \Omega} f(x_1, \ldots, x_n; \theta).$$

Note:

1. If each set of observations $(x_1, \ldots, x_n)$ corresponds to a unique value $\hat{\theta}$, then this procedure defines a function, $\hat{\theta} = t(x_1, \ldots, x_n)$. This same function when applied to the random sample, $\hat{\theta} = t(X_1, \ldots, X_n)$ is called

the **maximum likelihood estimator**, also denoted MLE.

2. Any value of $\hat{\theta}$ that maximizes $L(\theta)$ also will maximize the log-likelihood, $lnL(\theta) = l(\theta)$, so for computational convenience then alternate form of the maximum likelihood equation,
$$\frac{d}{d\theta}l(\theta)$$
often will be used.

Example 5. A binomial experiment consisting of n trials resulted in observations $x_1, x_2, \ldots, x_n$, where $x_i = 1$ if the i^{th} trial was a success and $x_i = 0$ otherwise. Find the MLE of p.

Example 6. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from an uniform distribution, $X_i \sim U(0, \theta)$. Find the MLE of θ.

Example 7.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $X_i \sim Beta(a = 6\theta, 1)$. Find the MLE of θ.

Example 8. Consider a random sample from two parameters exponential distribution, $X_i \sim Exp(1, \eta)$. Find the MLE of η.

Example 9. One observation is taken on a discrete random variable X with pdf $f(x|\theta)$, where $\theta = 1, 2, 3$. Find the MLE of θ.

x	$f(x\|1)$	$f(x\|2)$	$f(x\|3)$
0	$\frac{1}{3}$	$\frac{1}{4}$	0
1	$\frac{1}{3}$	$\frac{1}{4}$	0
2	0	$\frac{1}{4}$	$\frac{1}{4}$
3	$\frac{1}{6}$	0	$\frac{1}{4}$

Theorem 1. Invariance Property If $\hat{\theta}$ is the MLE of θ and if $u(\theta)$ is a function of θ, then $u(\hat{\theta})$ is an MLE of $u(\theta)$.

Example 10. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from an exponential distribution with parameter θ. Find the MLE of $P(X \geq 1)$.

The definitions of likelihood function and maximum likelihood estimator can be applied in the case of more than one unknown parameter if $\boldsymbol{\theta}$ represents a vector of parameters, say $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. if the partial derivatives of $L(\theta_1, \ldots, \theta_k)$ exist, and the MLEs do not occur on the boundary of Ω, then the MLEs will be solutions of the simultaneous equations

$$\frac{\partial}{\partial \theta_j} L(\theta_1, \ldots, \theta_k)$$

for $i = 1, \ldots, k$. These are called the **maximum likelihood equations** and the solutions are denoted by $\hat{\theta}_1, \ldots, \hat{\theta}_k$.

Theorem 2. Invariance Propefty If $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\ldots}, \hat{\theta}_k)$ denotes the MLE of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, then the MLE of $\boldsymbol{\tau} = (\tau_1(\theta), \ldots, \tau_k(\theta))$ is $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, ..., \hat{\tau}_k) = (\tau_1(\hat{\theta}), \ldots, \hat{\tau}(\hat{\tau}))$ for $= 1 \le r \le k$.

Example 11. For a set of random variables $X_i \sim N(\mu, \sigma^2)$, based on a random sample of size n, find the MLE of μ and σ^2 if both μ and σ^2 are unknown.

Example 12. For a set of random variables $X_i \sim GAM(\alpha, \theta)$, based on a random sample of size n, find the MLE of the mean.

4.3 Criteria For Evaluating Estimators

4.3.1 Unbiased Estimators

Definition 7. Unbiased Estimator An estimator T is said to be an unbiased estimator of $\tau(\theta)$ if

$$E(T) = \tau(\theta)$$

for all $\theta \in \Omega$. Otherwise, we say that T is a biased estimator of $\tau(\theta)$.

Example 13. Let $X_1, X_2, \ldots, X_n$ be a random sample with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$. Show that

(a) $S'^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$ is a biased estimator for σ^2 and that

(b) $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ is an unbiased estimator for σ^2.

It is possible to have a reasonable estimator that is biased, and often an estimator can be adjusted to make it unbiased

Example 14. Consider a random sample of size n from an exponential distribution, $X \sim EXP(\theta)$.

(a) Find the MLE of $\frac{1}{\theta}$, say T_1.

(b) Show that T_1 is a biased estimator of $\frac{1}{\theta}$.

(c) Find a constant c such that cT_1 is an unbiased estimator of $\frac{1}{\theta}$.

4.3.2 Mean Squared Error

Definition 8. If T is an estimator of $\tau(\theta)$, then the **bias** is given by

$$Bias(T) = E(T) - \tau(\theta)$$

and the **mean sqanred error (MSE)** of T is given by

$$MSE(T) = E[T - \tau(\theta)]^2 = E(T^2) - 2\tau(\theta)E(T) + \tau^2(\theta)$$

Theorem 3. If T is an estimator of $\tau(\theta)$, then

$$MSE(T) = V(T) + [Bias(T)]^2$$

Notes:

1. The MSE is a reasonable criterion that considers both the variance and the bias of an estimator, and it agrees with the variance criterion if attention is restricted to unbiased estimators.
2. It provides a useful means for comparing two or more estimators, but it is not possible to obtain an estimator that has uniformly minimum MSE for all $\theta \in \Omega$ and all possible estimators.

Example 15.

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a population whose density is given by

$$f(x) = \begin{cases} 3\beta^3 x^{-4}, & x \geq \beta \\ 0, & \text{otherwise} \end{cases}$$

where $\beta > 0$ is unknown. Consider the estimator $\hat{\beta} = X_{1:n}$. Derive the bias of the estimator.

Example 16.

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a population whose density is given by

$$f(x) = \begin{cases} 3\beta^3 x^{-4}, & x \geq \beta \\ 0, & \text{otherwise} \end{cases}$$

where $\beta > 0$ is unknown. Consider the estimator $\hat{\beta} = X_{1:n}$. Derive $MSE(\hat{\beta})$.

Example 17.

A random sample of size n is taken from a distribution with probability density funtion (pdf)

$$f(x) = \frac{5x^4}{\theta^5}, 0 < x < \theta, \text{zero otherwise.}$$

(a) Find the Maximum Likelihood Estimator(MLE) of θ. Call it $\hat{\theta}$.

(b) Find the Method of Moment Estimator(MME) of θ. Call it $\tilde{\theta}$.

(c) Find the Mean Square Error(MSE) of $\hat{\theta}$.

(d) Find the MSE of $\tilde{\theta}$.

4.3.3 Uniformly Minimum Variance Unbiased Estimators

Definition 9. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $f(x;\theta)$. An estimator T^* of $\tau(\theta)$ is called a uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$ if

1. T^* is unbiased for $\tau(\theta)$, and
2. for any other unbiased estimator T of $\tau(\theta)$,
$$V(T^*) \leq V(T)$$
for all $\theta \in \Omega$.

In some cases, lower bounds can be derived for the variance of unbiased estimators. If an unbiased estimator can be found that attains such a lower bound, then it follows that the estimator is a UMVUE.

Theorem 4. If T is an unbiased estimator of $\tau(\theta)$, then the **Cramer-Rao lower bound (CRLB**, based on a random sample, is

$$V(T) \geq \frac{[\tau'(\theta)]^2}{nE\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2}$$

Note: For T to attain the CRLB of $\tau(\theta)$, it must be a linear function of

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i;\theta).$$

Example 18.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $X_i \sim Beta(a = 6\theta, 1)$. Find the CRLB of θ.

Example 19.

Let $X_1, X_2, \ldots, X_n$ denote a random sample from the density function given by

$$f(x) = \begin{cases} \frac{5}{\theta} x^4 e^{-x^5/\theta}, & \theta > 0, x > 0, \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the MLE of θ.

(b) Find the CRLB of θ.

(c) Find the UMVUE for θ.

Example 20.

Let $Y_1, \ldots, Y_n$ be independent where $Y \sim N(\beta x_i, \sigma^2)$ with both β and σ^2 unknown.

(a) If $y_1, \ldots, y_n$ are observed, derive the MLEs $\hat{\beta}$ and $\hat{\sigma}^2$ based on the pairs $(x_1, y_1), \ldots, (x_n, y_n)$.

(b) Find the distribution of the estimator $\hat{\beta}$.

(c) Find c so that $c\hat{\sigma}^2$ is an unbiased estimator of σ^2.

(d) Find the CRLB of β.

4.3.4 Large-Sample Properties

Properties of estimators such as unbiasedness and uniformly minimum variance are defined for any fixed sample size n. These are examples of "small-sample" properties.

It also is useful to consider asymptotic or "large-sample" properties of a particular type of estimator. An estimator may have undesirable properties for small n, but still be a reasonable estimator in certain applications if it has good asymptotic properties as the sample size increases. It also is possible quite often to evaluate the asymptotic properties of an estimator when small sample properties are difficult to determine.

Definition 10. Simple Consistency Let $\{T_n\}$ be a sequence of estimators of $\tau(\theta)$. These estimators are said to be **consistent** estimators of $\tau(\theta)$ if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P[|T_n - \tau(\theta)| < \epsilon] = 1$$

for every $\theta \in \Omega$.

Definition 11. MSE Consistency If $\{T_n\}$ is a sequence of estimators of $\tau(\theta)$, then they are called **mean squared error consistent** if

$$\lim_{n\to\infty} E[T_n - \tau(\theta)]^2 = 0$$

for every $\theta \in \Omega$.

Definition 12. Asymptotic Unbiased A sequence $\{T_n\}$ is said to be asymptotically unbiased for $\tau(\theta)$ if

$$\lim_{n\to\infty} E[T_n - \tau(\theta)] = \tau(\theta)$$

for every $\theta \in \Omega$.

Theorem 5. A sequence $\{T_n\}$ of estimators of $\tau(\theta)$ is mean squared error consistent if and only if it is asymptotically unbiased and $\lim\limits_{n\to\infty} V(T_n) = 0$.

Theorem 6. If a sequence $\{T_n\}$ is mean squared error consistent, it also is simply consistent.

Example 21. Consider a random sample of size n from a distribution with pdf $f(x;\theta) = 1/\theta$ if $0 < x \leq \theta$, and zero otherwise; $0 < \theta$. Show that the MLE of θ is MSE consistent.

4.3.5 Asymptotic Properties of MLEs

Under certain circumstances, it can be shown that the MLEs have very desirable properties. Specifically, if certain regularity conditions are satisfied, then the solutions $\hat{\theta}_n$ of the maximum likelihood equations have the following properties

1. $\hat{\theta}_n$ exists and is unique,
2. $\hat{\theta}_n$ is a consistent estimator of θ,
3. $\hat{\theta}_n$ is asymptotically normai with asymptotic mean θ and variance $\frac{1}{n}E\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2$ and
4. $\hat{\theta}_n$ is asymptotically efficient.

Notes:

1. For large n, approximately
$$\hat{\theta} \sim N(0, CRLB \text{ of } \theta)$$
2. If $\tau(\theta)$ is a function with nonzero derivative, then $\hat{\tau}_n = \tau(\hat{\theta}_n)$ also is asymptotically normal with asymptotic mean $\tau(\theta)$ and variance $[\tau(\theta)]^2 CRLB$ of θ.

Example 22. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from an exponential distribution with parameter θ.Suppose that now we are interested in estimating

$$R = R(t;\theta) = P(X > t) = exp(-t/\theta)$$

What is the asymptotic distribution of R.

Example 23. Consider a random sample from a Pareto distribution, $X \sim PAR(\alpha, \theta(known))$.

(a) Find the MLE of α, $\hat{\alpha}$.

(b) Find the asymptotic distribution of $\hat{\alpha}$.

4.4 Bayes Estimation

The Bayesian statistician assumes that the universe follow a parametric model, with unknown parameters. The distribution of the model given the value of the parameters is called the model distribution. Unlike frequentist, who estimates the parameters from the data, the Bayesian assigns a prior probability distribution to the parameters. After observing data, a new distribution, called the posterior distribution is developed for the parameters.

4.4.1 Posterior Distribution

Definition 13. Posterior Distribution
The conditional density of θ given the sample observations $\mathbf{x} = (x_1, \ldots, x_n)$ is called the posterior density or posterior pdf, and is given by

$$\pi_{\theta|\mathbf{x}}(\theta) = \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{\int f(x_1, \ldots, x_n|\theta)\pi(\theta)} d\theta = \frac{f(\mathbf{x}|\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)} d\theta$$

Posterior Mean, $E(\Theta) = \int \theta \pi_{\theta|\mathbf{x}}(\theta) d\theta = E[E(\mathbf{X}|\theta)]$

4.4.2 Estimation

Definition 14. Loss Function If T is an estimator of $\tau(\theta)$, then a loss function is any real-valued function, $L(t;\theta)$, such that

$$L(t;\theta) \geq 0 \text{ for every } t$$

and

$$L(t;\theta) = 0 \text{ when } t = \tau(\theta)$$

Definition 15. If $X_1, \ldots . X_n$ denotes a random sample from $f(x|\theta)$, then the Bayes estimator is the estimator that minimizes the expected loss relative to the posterior distribution, $\theta|\mathbf{x}$,

$$E_\theta[L(T;\theta]$$

Three loss function that are commonly use by Bayesian are:

1. Square error loss function, $L(T;\theta) = [T - \theta]^2$
2. Absolute error loss function, $L(T;\theta) = |T - \tau(\theta)|$
3. Almost constant loss function,

$$L(T;\theta) = \begin{cases} c, T \neq \theta, \\ 0, \text{otherwise} \end{cases}$$

Theorem 7.
The Bayes estimator, T, of θ under the squared error loss function,

$$L(T;\theta) = [T-\theta]^2$$

is the conditional mean of θ relative to the posterior distribution,

$$T = E_{\theta|\mathbf{x}}[\theta] = \int \theta f_{\theta|\mathbf{x}}(\theta) d\theta$$

Theorem 8. The Bayes estimator, T, of θ under absolute error loss,

$$L(T;\theta) = |T-\theta|$$

is the median of θ relative to the posterior distribution.

Theorem 9.
If the loss function is $L(T;\theta) = \begin{cases} c, \hat{\theta} \neq \theta \\ 0, \text{otherwise} \end{cases}$,then the Bayesian point estimator, $\hat{\theta}$, which minimizes the expected value of the loss function is the mode of the posterior distribution.

Example 24.
Suppose

$$f(x_i|\theta) = \frac{4\theta^4}{x_i^5}, x_i \geq \theta$$

and

$$\pi(\theta) = \frac{1}{10}, 3 \leq \theta \leq 13.$$

3 observations: 5, 7, and 10 were observed. Calculate the Bayesian point estimate using the 3 loss function just described.

Example 25.

Suppose $X|\theta \sim U(\theta-\frac{1}{7}, \theta+\frac{6}{7})$ and that a prior distribution of θ is $N(0,1)$. Find the Bayes estimator of θ under squared error loss and Bayes estimate when $x = 0.82$.